ESO #446

9.04
446

A FORTRAN COMPUTER PROGRAM FOR
CALCULATING THE GINI RATIO FOR
UNGROUPED DATA

By

Marcia M. Gowen
Linda Buttel
Richard L. Meyer

Department of Agricultural Economics and Rural Sociology
The Ohio State University
Columbus, Ohio
January 3, 1978

A FORTRAN COMPUTER PROGRAM FOR CALCULATING THE GINI RATIO OF UNGROUPED DATA

Introduction

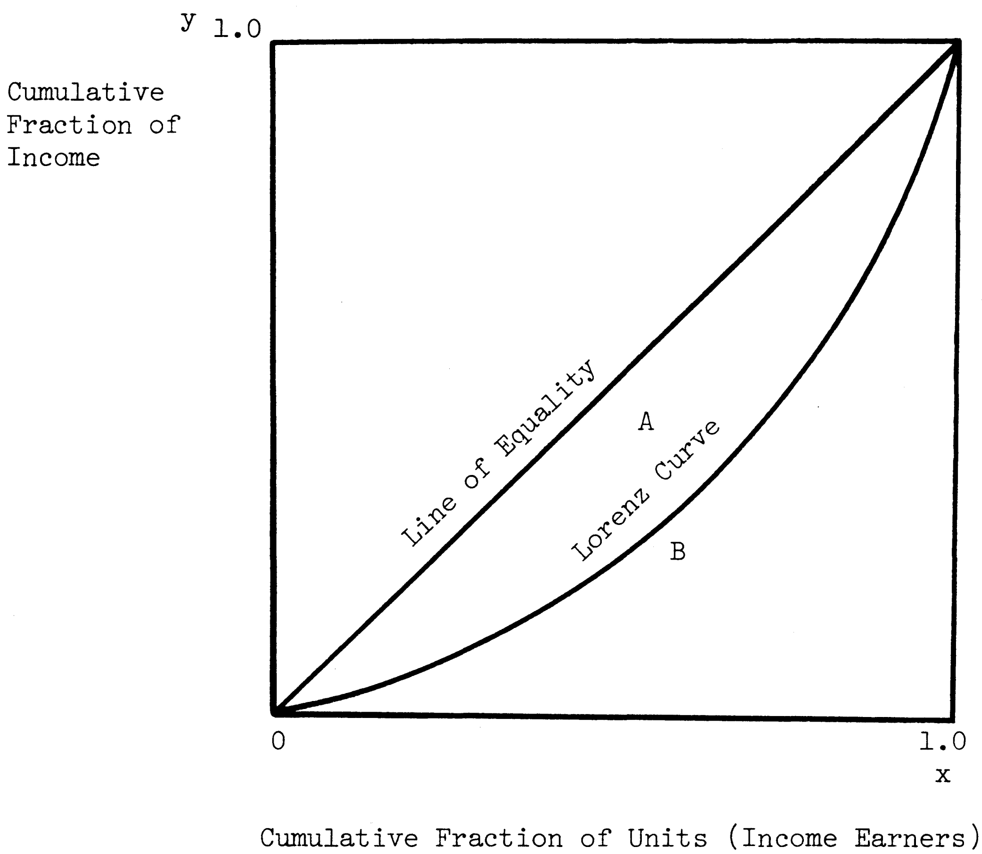
Recent interest in income distribution has encouraged the creation of alternative measures of income inequality. The Gini Ratio, or Gini Index of Concentration, is one commonly accepted measure of income inequality. This paper presents (1) how the index is calculated based on theory and (2) a computer program designed to calculate the Gini Ratio for ungrouped data. Though vital to a full understanding of the applicability of this index, a discussion of the limitations of this index is not included but may be found elsewhere.[1/] This paper is limited to a discussion of the concepts involved in calculating the ratio and a Fortran computer program for the Gini for income inequality using ungrouped data. The example shown is for rural household income distribution in Taiwan. The approach would be similar if the distribution of another variable, say, farm area, was desired.

The Calculation of the Gini Ratio

The Gini Ratio represents the proportion of the triangular area in a unit square falling below the Lorenz curve. Therefore, to conceptually understand the Gini Ratio, the Lorenz curve must first be understood.

A Lorenz curve may be derived by plotting the cumulative fraction of units (income earners in the case reported in this paper) arrayed in order from the smallest to the largest income (the X-axis) against the cumulative share of the aggregate income accounted for by these units (Y-axis). Within a unit square, a 45° diagonal line is drawn, known as the Line of Equality (Figure 1). Perfect equality of incomes among all units or income earners would result in such a line. Similarily if each income group's or percentile's income share of the total income exactly equaled their proportion of the population such a line would exist.

Figure 1. Lorenz Curve



Inequality of income among units or the existence of income groups not earning exactly their appropriate proportion of total income results in the Lorenz curve falling below the Line of Equality. The less the inequality, the closer the Lorenz curve falls relative to the Line of Equality. Thus Region A shown in Figure 1 is smallest when income equality is greatest, and as will be shown later, the Gini Ratio is zero if perfect equality exists. Conversely, the greater the inequality, the further the Lorenz curve lies from the Line of Equality.[2/]

The Gini Ratio is the proportion of area between the Line of Equality and the Lorenz curve divided by the total area under the Line of Equality:

$$(1)\ \text{GINI RATIO} = \frac{A}{A+B} = \frac{\text{Area Between Diagonal and Curve}}{\text{Total Area Under Diagonal}}$$

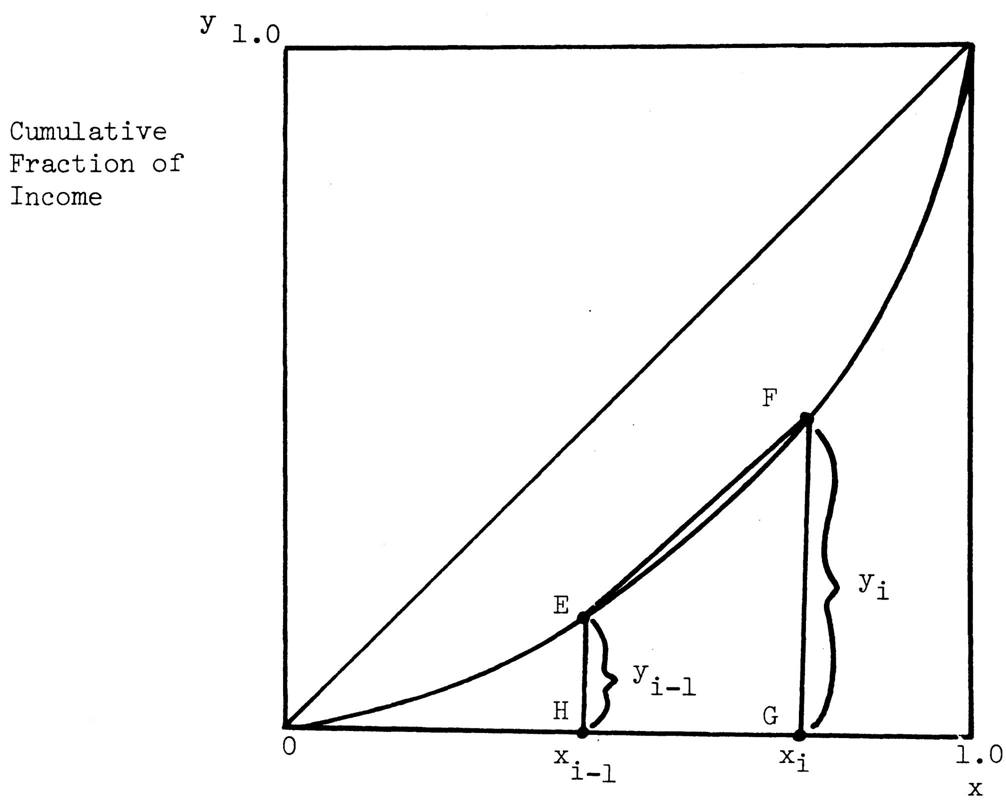
Since the figure is a unit square, the area under the diagonal equals one-half. Thus equation 1 can be rewritten as follows:

$$(2)\ \text{GINI RATIO} = \frac{1/2 - \text{Area B}}{1/2} = 1 - 2\ (\text{Area B})$$

Data for calculating the Gini Ratio may be either grouped into income percentiles or ungrouped. Different methods to estimate a Gini Ratio exist for the two types of data. Grouped data frequently exist in land distribution. For example, farms are sorted into arbitrary size classes and the cumulative number of farms and proportions of area reported.

A linear approximation of the Lorenz curve is used for calculating the Gini Ratio from grouped data. The calculation estimates the area under the curve by drawing straight lines between data points ($\overline{EF}$ in Figure 2), taking the area of each resulting polygon (EFGH), and summing the areas of the

Figure 2. Lorenz Curve



Cumulative Fraction of Units (Income Earners) several polygons to approximate the area under the curve. Because the straight lines connecting these data points lie above the curve, a Gini Ratio results which underestimates the true concentration index. Obviously, the greater the number of polygons created from a data set -- i.e., the greater the number of groups or percentiles -- the closer will be the estimate of the areas to the true area.[3/]

The calculation of the Gini Ratio using the linear approximation method can be expressed as follows:

Area under any line segment equals:

(3) Area EFGH = $\frac{(Y_i + Y_{i-1})}{2} (X_i - X_{i-1})$

Where: Y_i = Cumulative fraction of income

X_i = Cumulative fraction of units (income earners)

Summing over all the intervals to approximate the area under the curve gives:

$$(4) \quad S_i = \sum_{i=1}^{k} \frac{(Y_i + Y_{i-1})}{2} (X_i - X_{i-1})$$

Where: K = the number of intervals.

Through substitution of equation (4) into equation (2), a formula for estimating the Gini Ratio results:

$$(5) \quad \text{GINI RATIO} = 1 - 2 \sum_{i=1}^{k} \frac{(Y_i + Y_{i-1})}{2} (X_i - X_{i-1})$$

$$= 1 - \sum_{i-1}^{k} (Y_i + Y_{i-1}) (X_i - X_{i-1})$$

Similar presentations are given by Riemenschneider, Morgan, Bonnen, and Monke. Also, Gastwirth gives a somewhat different, though, in essence parallel presentation.

Several methods exist for calculating the Gini Ratio from ungrouped data. One such method uses the cumulative number of recipients (W_i) and the cumulative income (Z_i). The equation is as follows:

$$(6)\ \text{GINI RATIO} = 1 - \sum_{i=1}^{N} \frac{(Z_i + Z_{i-1})(W_i - W_{i-1})}{W_N \cdot Z_N}$$

Where: $i = 1, 2, 3 \ldots N$

$= \#$ of recipients

Other methods for determining the Gini Ratio from ungrouped data are given by Riemenschneider.

A Fortran Program for Calculating the Gini Ratio

This computer program calculates the Gini Ratio from ungrouped data using equation (5) instead of equation (6) because income data is most commonly expressed as income shares for income groups. The program (1) takes unordered, ungrouped observations, (2) orders them from lowest to highest income level, (3) divides the resulting list into separate groups representing income groups or percentiles (X_i's), (4) calculates the cumulative income for each group, (5) divides this cumulative group income by the number of observations in each group to obtain a mean group income, often called the income share (Y_i's), then (6) calculates the Gini Ratio using equation (5).

Several characteristics of the program are unique to this study and need elaboration. First, the number of observations varies from one year to the next for the data analyzed with this program. It was necessary to ensure that no observations would be dropped when dividing them into income groups. The program, thus, groups observations by taking the cumulative number of observations for one cumulative percentile and subtracts from it the number of observations in the previous cumulative percentile. For instance, to obtain the number of observations in the 51-60 percent group, the first 60 percent of the total number of observations in a given year is determined, then from it is subtracted the cumulative number of observations in the last 50 percent of the sample, i.e., number of observations in 51-60 percent = (.60N) - (.50N) where N = total number of observations.

Some control cards for the program will need adjustment depending on the specific data. They include the year of the sample, and the specification of income groups. The cards which pertain to these characteristics are designated on the program deck printout. Missing data is not accounted for in the program.

The printout of the program deck and parts of one program printout follow. The input comes from unit 11 and output is on unit 6 (which is universally the printer). Brief explanations of the program results are found on the printout.

Printout of Computer Program

*those items possibly to be changed

```
C                                  THIS PROGRAM CALCULATES THE GINI COEFFICIENT
C                                  FOR UNGROUPED DATA
C                                  IT ASSUMES THE DATA HAS BEEN PRESORTED IN
C                                  ASCENDING ORDER ON THE VARIABLE BEING CONSIDERED
    * REAL*8 TOTIN(15),SHARE(15),X(15),CUMSH(15),GINI,PER,IPER(15)
    * DIMENSION  IRANK(1100),NUM(15)
    * DATA X/9*.1,2*.05/,INT/12/
C                                  INITIALIZE THE GINI COEFFICIENT
      DATA GINI/0./,PER/0./
C                                  INITIALIZE TOTAL INCOME
    * DATA TOTIN/15*0./
C                                  DEFINE YEAR, VARIABLE AND NUMBER OF CASES IN
    * DATA IYR/1973/,JVAR/173/,ICASES/460/
C                                  THE SAMPLE FOR THIS RUN
C                                  SET UP PERCENTILE INTERVALS AND NUMBER OF
C                                  INTERVALS,INCLUDING THE TOTAL
      CASES=ICASES
      DO 60 J=1,INT
      IF(J.EQ.INT) GO TO 60
      PER=PER+X(J)
      IPER(J)= CASES*PER +0.5
      NUM(J)=IPER(J)
      K=1
      L=IPER(J)
      IF(J.EQ.1) GO TO 62
      K=IPER(J-1)+1
      NUM(J)=L-K+1
   62 DO 61 I=K,L
      IRANK(I)=J
   61 CONTINUE
   60 CONTINUE
      NUM(INT)=ICASES
      I=0
   90 READ(11,100,END=10) IVAR
      I=I+1
C                                  SET UP PERCENTITLE
      J=IRANK(I)
C                                  CALCULATE INCOME FOR A GIVEN PERCENTILE
      TOTIN(J)=TOTIN(J)+IVAR
C                                  CALCULATE TOTAL INCOME FOR SAMPLE
      TOTIN(INT)=TOTIN(INT)+IVAR
      GO TO 90
   10 CONTINUE
C                                  CALCULATE INCOME SHARES
      DO 80 J=1,INT
      SHARE(J)=TOTIN(J)/TOTIN(INT)
      IF(J.EQ.1) GO TO 81
C                                  CALCULATE CUMULATIVE SHARES
      CUMSH(J)=CUMSH(J-1)+SHARE(J)
      GO TO 80
   81 CUMSH(1)=SHARE(1)
   80 CONTINUE
      CUMSH(INT)=CUMSH(INT-1)
      WRITE(6,200) JVAR,IYR
      WRITE(6,201)
      WRITE(6,202) (NUM(J), J=1,INT),(TOTIN(J),J=1,INT),(SHARE(J),J=1,IN
     +T),(CUMSH(J),J=1,INT)
C                                  CALCULATE THE GINI COEFFICIENT
      DO 70 J=1,INT
      IF(J.EQ.1) GO TO 71
      IF(J.EQ.INT) GO TO 70
      GINI=GINI+((CUMSH(J)+CUMSH(J-1))*X(J))
```

(Continued)

```
      GO TO 70
   71 GINI=GINI+(CUMSH(J)*X(J))
   70 CONTINUE
      GINI=1.-GINI
      WRITE(6,300) GINI
  100 FORMAT(250X,250X,250X,250X,177X,I7)
  200 FORMAT('1',45X,'VARIABLE = ',I3,10X,'YEAR = ',I4/'0')
  201 FORMAT('0',24X,' 1-10',4X,'11-20',4X,'21-30',4X,'31-40',4X,'41-50'
     +,4X,'51-60',4X,'61-70',4X,'71-80',4X,'81-90',4X,'91-95',3X,'96-100
     +',7X,'TOTAL'/' ')
  202 FORMAT('ONUMBER OF CASES ',5X,11(I8,1X),2X,I9/'OINCOME TOTAL',9X,1
     +1(F8.0,1X),2X,F9.0/
     +       'OINCOME SHARE',9X,11(F8.5,1X),2X,F9.5/'OCUMULATIVE SHARE',
     +5X,
     +11(F8.5,1X),2X,F9.5/'0')
  300 FORMAT('OTHE GINI COEFFICIENT IS ',F10.5)
      STOP
```

Printout of Program Results

Percentiles

VARIABLE = 173   YEAR = 1973

	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-95	96-100	TOTAL
NUMBER OF CASES	46	46	46	46	46	46	46	46	46	23	23	460
INCOME TOTAL	1697288.	2403743.	2797990.	3171335.	3601136.	4044426.	4570294.	5362812.	6464784.	4022215.	[illegible]	44065521.
INCOME SHARE	0.03852	0.05455	0.06350	0.07197	0.08172	0.09178	0.10372	0.12170	0.14671	0.09128	0.13456	1.00000
CUMULATIVE SHARE	0.03852	0.09307	0.15656	0.22853	0.31025	0.40204	0.50575	0.62745	0.77416	0.86544	1.00000	1.00000

THE GINI COEFFIEIENT IS 0.27490

Footnotes

1. A thorough discussion of the limitations of the Gini Ratio may be found in Riemenschneider. Also Eric Monke makes a comparison of different income measures. Paglin modifies the calculations in an attempt to take into account the life-cycle effects on an income distribution from a sample population which contains income earners of various ages.

2. It is obvious that the Gini Ratio does not give complete information concerning the distribution of the variable. Other statistics such as the mean, mode, median, coefficient of skewness, Kurtosis, and the coefficient of variation provide further illustrative information to describe the distribution.

3. Riemenschneider sets the minimum number of percentile groups at eight to obtain the closest approximation of the true area. Unfortunately most income data from countries exist in groups ranging from the lowest twenty percent of income earners to the top five percent (0-20 percent, 21-40 percent, 41-60 percent, 61-80 percent, 81-95 percent, 96-100 percent). Felix Paukeit's controversial study on income distribution (1973) based on an extensive data collection helped to set this pattern. To reduce possible underestimates of the Gini Ratio it is advisable to have at least eight groups.

References

Bonner, James T., "The Distribution of Benefits from Selected U.S. Farm Programs", Rural Poverty in the United States, A report by the President's National Advisory Commission on Rural Poverty, Washington, D. C., May, 1968.

Gastwirth, J. L., "The Estimation of the Lorenz Curve and Gini Index", The Review of Economics and Statistics, Vol. 54, No. 3, August, 1972, pp. 306-316.

Monke, Eric, "Income Distribution in Taiwan: A Panel Study of Agricultural Farm Families", unpublished paper, Food Research Institute, Stanford University, August 1, 1975.

Morgan, James, "The Anatomy of Income Distribution", Review of Economics and Statistics, Vol. 44, No. 3, August, 1962, pp. 207-283.

Paukert, Felix, "Income Distribution at Different Levels of Development: A Survey of Evidence", International Labour Review, August-September, 1973, pp. 97-125.

Reimenschneider, Charles, "The Use of the Gini Ratio in Measuring Distributional Impacts", unpublished paper, Department of Agricultural Economics, Michigan State University, circa 1976.